A shock-tracking algorithm for surface evolution under reactive-ion etching

Cite as: Journal of Applied Physics **74**, 5172 (1993); https://doi.org/10.1063/1.354282
Submitted: 01 September 1992 . Accepted: 04 July 1993 . Published Online: 04 June 1998

S. Hamaguchi, M. Dalvie, R. T. Farouki, and S. Sethuraman

A shock-tracking algorithm for surface evolution under reactive-ion etching

S. Hamaguchi, M. Dalvie, R. T. Farouki, and S. Sethuraman[a)]
IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598

(Received 1 September 1992; accepted for publication 4 July 1993)

A new algorithm that determines the evolution of a surface eroding under reactive-ion etching is presented. The surface motion is governed by both the Hamilton–Jacobi equation and the entropy condition for a given etch rate. The trajectories of "shocks" and "rarefaction waves" are then directly tracked, and thus this method may be regarded as a generalization of the method of characteristics. This allows slope discontinuities to be accurately calculated without artificial diffusion. The algorithm is compared with "geometric" surface evolution methods, such as the line-segment method.

I. INTRODUCTION

An understanding of the evolution of microscopic surface features under etching or deposition processes is of considerable practical importance in integrated-circuit manufacturing. Recent progress in semiconductor fabrication technologies has given rise to complex manufacturing processes, and a correspondingly increased demand for computational models of surface-morphology evolution that facilitate analysis of the effects of variations in process-step parameters. In this paper, a new algorithm for evolving surfaces is presented. The algorithm is derived from first principles, based on mathematical theory that describes a well-defined physical system.

To date, a number of surface-evolution methods have been proposed for microelectronics manufacturing processes. These techniques may be broadly categorized along the following lines.

(1) "Geometric" models—the surface evolution is governed by *ad hoc* geometric considerations. Thus, the surface is generally discretized into a set of "nodes" connected by linear segments. Depending on the assumed rules for how the nodes and/or segments are moved, such methods are called "string models," "line-segment models,"..., etc. Although they appear capable of generating surface morphologies that are consistent with empirical micrographic cross sections, such models are fundamentally based on the implementer's geometric intuition rather than a well-defined and self-consistent mathematical theory (even for idealized physical conditions). Thus, such models suffer from a basic lack of credibility in terms of *predictive* capacity, especially with regard to the formation and evolution of "singular" features, such as sharp corners.

(2) Method of characteristics—as shown in Sec. II, a moving surface may be described by a partial differential equation (PDE). If the surface velocity (i.e., the etch or deposition rate) does not depend on the local surface curvature, or higher-order intrinsic geometry, then the PDE is of first order and can be solved by the method of characteristics. Problems arise, however, when discontinuities of the surface gradient form (i.e., two characteristics intersect). Existing models based on the method of characteristics of which we are aware appear to require subtle geometric "adjustments," such as eliminating or avoiding the formation of nonphysical loops—"delooping"—based on geometric hypotheses to deal with such situations (see, for example, Refs. 1 and 2).

(3) PDE solvers—there are a number of methods to solve the PDE directly on a (fixed) spatial grid, without appealing to the characteristic equations. Such methods are particularly appropriate when the surface velocity depends explicitly on the local curvature and the PDE is of second order.[3–5] The second-order term, which typically originates from the curvature-dependence of the velocity, introduces a "diffusion" effect so that slope discontinuities are not formed in this case. The solution in the limit of zero diffusion, known as the "viscosity solution," is the desired solution for systems in which the surface velocity does not depend on the local curvature. PDE solvers using a combination of a Lax–Wendroff-type finite-difference scheme with an upwind (or artificial diffusion) scheme rely on a small "numerical" diffusion to obtain an approximation to the viscosity solution.

The method proposed in this paper employs nodal points connected by linear segments to represent the evolving surface, as in the "geometric" models mentioned above. However, unlike the geometric models, the motion of the nodes and segments is derived from a PDE and an additional condition that together govern the surface evolution. In fact, the use of linear segments between nodes is merely a consequence of the lowest-order spatial discretization of the system, and extending the method to higher-order schemes employing polynomial curve segments should be possible. At present we shall treat only the lowest-order scheme, so as to focus our attention on deriving the surface dynamics from the defining equations.

Our method may be regarded as an extension of the method of characteristics, although explicit solutions of the characteristic equations are not employed by the method. Instead, we solve for the propagation of "shocks" (see Sec. II below) that emanate from each nodal point, assuming

[a)]Permanent address: Courant Institute of Mathematical Sciences, New York University, New York, NY 10012.

 0021-8979/93/74(8)/5172/13/$6.00

that each node connecting adjacent linear segments represents a surface-slope discontinuity. It will be shown that such shock trajectories are good approximations of characteristics under certain conditions.

Another feature of our method that differentiates it from the usual method of characteristics is the implementation of the entropy condition, which selects the physically correct solution of the surface-evolution equation. As will be explained in detail below, the PDE that describes the surface evolution—such as Eq. (2) or (3)—generally has no unique solution, and thus an extra condition must be imposed to identify the physically meaningful solution. This additional condition is called the entropy condition, and is proven to select the viscosity solution mentioned above. Thus, the direct imposition of the entropy condition enables one to identify the unique physical solution, without invoking *ad hoc* geometric procedures to contend with nonphysical contingencies such as "loop" formation.

Since the shock (i.e., facet corner) trajectories are tracked explicitly, our method may be classified as a "shock-tracking" method. (Several different types of shock-tracking method exist to solve systems of conservation laws: see, for example, Refs. 6 and 7.) Under certain conditions, the motion of a surface described by this shock-tracking method may be shown to agree with that obtained from "geometric" methods. Therefore, the shock-tracking method may also be regarded as offering a theoretical foundation for some of the "intuitive" assumptions made in geometric methods.

For the sake of simplicity, we consider only the evolution of "one-dimensional" surfaces (i.e., curves in the x-y plane) in this paper. Extending the shock-tracking method to surfaces in three dimensions should be possible, and will be addressed in a future paper. Furthermore, so as to focus on the numerical algorithm of advancing such a curve with time, we shall idealize all physical/chemical details of the etch process, assuming they are adequately represented by a simple rate function c that describes the surface velocity under unidirectional ion bombardment. Important physical and chemical aspects of the problem, such as the velocity distribution and shadowing (visibility angle) of the incoming ion flux,[8,9] reflection, and deposition of the incoming ions and sputtered materials, and diffusion and chemical reactions on the surface, are deferred to future studies that will make use of the basic algorithm described below.

II. SURFACE EVOLUTION EQUATIONS

We first derive the equation of motion for the boundary surface that separates regions of material and free space. In two dimensions, such a boundary surface corresponds to a piecewise-smooth curve that does not intersect itself. At time t, such a curve may be represented by an equation of the form $\phi(x,y,t)=0$, where we assume that $\phi(x,y,t)>0$ (<0) represents the material (vacuum) side of the boundary. The velocity vector $\mathbf{c}$ of the boundary surface at the point (x,y) may be written as

$$\mathbf{c}=c_n\hat{\mathbf{n}}+c_t\hat{\mathbf{t}}, \tag{1}$$

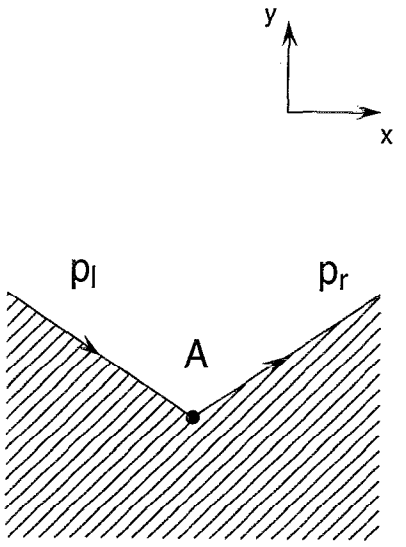

FIG. 1. A slope (i.e., gradient) discontinuity at point A. The slopes on the left and right sides of point A are given by p_l and p_r. The shaded area represents the material. The arrows on the boundary curve denote the orientation.

where $\hat{\mathbf{n}}$ and $\hat{\mathbf{t}}$ denote the unit normal and tangent unit vectors, defined by

$$\hat{\mathbf{n}}=\frac{1}{\sqrt{\phi_x^2+\phi_y^2}}\begin{pmatrix}\phi_x\\ \phi_y\end{pmatrix},\quad \hat{\mathbf{t}}=\frac{1}{\sqrt{\phi_x^2+\phi_y^2}}\begin{pmatrix}-\phi_y\\ \phi_x\end{pmatrix},$$

and $\phi_x=\partial\phi/\partial x$ and $\phi_y=\partial\phi/\partial y$. Note that, for $c_n>0$, the boundary curve moves into the material, i.e., its motion represents the erosion (or etching) of material.

Introducing a parametric variable s along the boundary curve, we may represent this curve at each time t as the locus of points $(x(s,t),y(s,t))$ for $0\leqslant s\leqslant 1$, where the functions x and y are assumed to be continuous and piecewise differentiable with respect to s, and the velocity of the point $(x(s,t),y(s,t))$ with fixed s is given by Eq. (1). Clearly, $\phi(x(s,t),y(s,t),t)\equiv 0$. We may also choose the orientation of the curve—i.e., the direction of increasing s—in such a way that the material (vacuum) always lies to the right (left) of the curve (see Fig. 1).

Differentiating the equation $\phi(x(s,t),y(s,t),t)=0$ with respect to time t (with s fixed) and substituting the relation $(\partial x/\partial t,\partial y/\partial t)=\mathbf{c}$ given by Eq. (1) into the resulting equation, we obtain

$$\phi_t+c\sqrt{\phi_x^2+\phi_y^2}=0, \tag{2}$$

where we write $c=c_n$. If the normal-direction etch rate c depends only on time t, position (x,y), the unknown function ϕ and its first derivatives (ϕ_x,ϕ_y), Eq. (2) is called a Hamilton–Jacobi equation. Equation (2), together with the initial condition $\phi(x,y,0)$, governs the evolution of the boundary curve. Note that the tangential velocity component (etch rate) c_t does not appear in Eq. (2). The velocity along the curve does not alter its shape, and the motion of the curve is determined only by the normal component $c=c_n$.[10]

If the "height" of the boundary curve y is uniquely determined by the horizontal position x, i.e., the curve at time t is given by a single-valued function $y=u(x,t)$, we may write $\phi(x,y,t)=u(x,t)-y$, and Eq. (2) becomes

$$u_t+c\sqrt{1+u_x^2}=0. \tag{3}$$

It is now assumed that $c=c(t,x,y;u_x)$. By differentiating both sides of Eq. (3) and writing $p=u_x$,[3,4] we obtain the conservation form

$$p_t+\partial_x f(t,x,y,p)=0, \tag{4}$$

where $f(t,x,y,p)=c\sqrt{1+p^2}$ is called the "flux function" by analogy[11] with the conservation law in gas dynamics. We note that f is proportional to the sputtering yield (the number of sputtered atoms per incident ion) in the case of inert ion beam etching.

III. VISCOSITY SOLUTIONS

We now briefly discuss the viscosity solutions of the Hamilton–Jacobi equation. Adding a "diffusion" term to Eqs. (2), we obtain

$$\phi_t^{(\nu)}+c\sqrt{\phi_x^{(\nu)2}+\phi_y^{(\nu)2}}=\nu\Delta\phi^{(\nu)}, \tag{5}$$

where $\Delta=\partial^2/\partial x^2+\partial^2/\partial y^2$ denotes the Laplacian, and we call the constant ν the diffusion constant. When ν is small, we may regard the right-hand side of Eq. (5) as an approximation of the effect of weak dependence of the etch/deposition rate on the local curvature K, i.e., etch/depostion rate$=c-\nu K$.[5] This weak curvature dependence of the etch/deposition rate results in "rounding off" of sharp corners of the surface. It is known that, under appropriate initial and boundary conditions, the smooth solution $\phi^{(\nu)}$ to Eq. (5) exists uniquely if $\nu\neq 0$. Since we may view Eq. (2) as the $\nu\to 0$ limit of Eq. (5), we expect that the physically meaningful solution to Eq. (2) should be the limit of the unique solution $\phi^{(\nu)}$ to Eq. (5) as $\nu\to 0$.

The limiting solution $\phi^{(0)}\equiv\lim_{\nu\to 0}\phi^{(\nu)}$ may contain some sharp corners, i.e., its spatial derivatives may not exist at certain points, due to the absence of the diffusion effect. Nevertheless, such a solution [which is the physically meaningful one that we wish to obtain from Eq. (2)], will satisfy Eq. (2), except at points where ϕ is not differentiable (the equation is meaningless if ϕ_x and ϕ_y are not defined). Therefore, we relax the definition of the solution of Eq. (2) by calling the function ϕ a solution if ϕ satisfies Eq. (2) everywhere except for those points where ϕ is not differentiable (i.e., the locations of sharp corners). Such solutions with discontinuous gradients are called "weak" or "generalized" solutions, in contrast to the regular (i.e., differentiable) solutions.

However, there are actually many different weak solutions to Eq. (2), only one of which is physically meaningful (a manifestation of such nonphysical solutions is the phenomenon of loop formation, observed in "geometric" models and the method of characteristics—see, for example, Refs. 12 and 13). In other words, Eq. (2) does not on its own constitute a complete description of the physical surface evolution: we need an additional condition that selects the physical solution from among the many weak solutions admitted by Eq. (2). This condition, which is called the entropy condition, will be discussed in Sec. VI.

The viscosity solution of Eq. (3) may be defined in a similar manner. Consider the hyperbolic equation

$$u_t^{(\nu)}+c\sqrt{1+u_x^{(\nu)2}}=\nu u_{xx}^{(\nu)}, \tag{6}$$

where ν again denotes the diffusion constant. By differentiating both sides of Eq. (6), we obtain

$$p_t^{(\nu)}+\partial_x f(t,x,y,p^{(\nu)})=\nu p_{xx}^{(\nu)}. \tag{7}$$

The viscosity solutions to Eqs. (3) and (4) are defined as $\lim_{\nu\to 0}u^{(\nu)}$ and $\lim_{\nu\to 0}p^{(\nu)}$, respectively. Further discussion of these viscosity solutions may be found in Ref. 3.

IV. THE CHARACTERISTIC EQUATIONS

For simplicity, we first consider Eq. (3), where the boundary curve is assumed to be describable by a function $y=u(x,t)$. The discussion will be later extended to the general case of Eq. (2). The initial value problem for the first-order PDE (3) may be solved by the method of characteristics as long as the curve does not intersect itself and remains sufficiently smooth (see, for example, Ref. 14). If the rate function c depends only on time, position on the curve, and its slope, i.e., $c=c(t,x,y;p)=c(t,x,u;u_x)$, as before, the characteristic equations of PDE (3) are given by

$$\frac{dx}{dt}=f_p=c_p\sqrt{1+p^2}+c\frac{p}{\sqrt{1+p^2}}, \tag{8}$$

$$\frac{dy}{dt}=pf_p-f=c_p p\sqrt{1+p^2}-c\frac{1}{\sqrt{1+p^2}}, \tag{9}$$

$$\frac{dp}{dt}=-f_x-pf_u=-(c_x+pc_u)\sqrt{1+p^2}, \tag{10}$$

where the subscripts denote partial derivatives, e.g., $c_p=\partial c/\partial p$. The initial values for Eqs. (8)–(10) are given by a point (x_0,y_0,p_0) on the initial curve $y=u(x,0)$, where $y_0=u(x_0,0)$ and $p_0=u_x(x_0,0)$. Thus, the initial-value problem for the first-order PDE (3) has been transformed into the system of ordinary differential equations (8)–(10).

If we employ the angle variable θ defined by $p=\tan\theta$, with $-\pi/2\leqslant\theta\leqslant\pi/2$, Eqs. (8)–(10) become

$$\frac{dx}{dt}=c_\theta\cos\theta+c\sin\theta, \tag{11}$$

$$\frac{dy}{dt}=c_\theta\sin\theta-c\cos\theta, \tag{12}$$

$$\frac{d\theta}{dt}=-c_x\cos\theta-c_u\sin\theta, \tag{13}$$

where we write $c(t,x,y;\tan\theta)=c(t,x,y,\theta)$ and $c_\theta=\partial c/\partial\theta=c_p/\cos^2\theta$. It is clear from these equations that, if $c_\theta\neq 0$, the direction of a characteristic curve is not perpendicular to the boundary curve $y=u(x,t)$. We also note that if the rate function c depends only on the slope (i.e.,

$c_x=c_u=0$), then the slope $p=\tan\theta$ of the boundary curve remains constant along every characteristic, and thus the characteristic curves become straight lines.

It is straightforward to extend the characteristic equations obtained above to the general case described by Eq. (2). If $\phi_y\neq 0$, Eq. (2) may be written as

$$\frac{\phi_t}{\phi_y}+c\ \mathrm{sgn}\ \phi_y\sqrt{1+\left(\frac{\phi_x}{\phi_y}\right)^2}=0, \tag{14}$$

where $\mathrm{sgn}\ \phi_y=1$ if $\phi_y>0$ and $\mathrm{sgn}\ \phi_y=-1$ if $\phi_y<0$. Note that, in the case of $\phi(x,y,t)=u(x,t)-y$, we have $u_t=-\phi_t/\phi_y$ and $p=u_x=-\phi_x/\phi_y$ and Eq. (14) is equivalent to Eq. (3). In general, we employ the definition of the extended angle variable θ $(-\pi<\theta\leqslant\pi)$ given by

$$\tan\theta=-\phi_x/\phi_y,$$

with

$$-\pi<\theta\leqslant-\pi/2 \quad \text{or} \quad \pi/2\leqslant\theta\leqslant\pi, \quad \text{if } \phi_y\geqslant 0,$$
$$-\pi/2\leqslant\theta\leqslant\pi/2, \quad \text{if } \phi_y\leqslant 0. \tag{15}$$

By comparing Eq. (14) with Eq. (3), we may also extend the definition of the flux function f to

$$f=-c\ \mathrm{sgn}\ \phi_y\sqrt{1+\left(\frac{\phi_x}{\phi_y}\right)^2}=\frac{c}{\cos\theta}. \tag{16}$$

If the rate function c in Eq. (2) has the form $c=c(t,x,y;-\phi_x/\phi_y)$, then it is easy to show that the characteristic equations for Eq. (2) may be reduced to Eqs. (11)–(13) with the extended angle θ $(-\pi<\theta\leqslant\pi)$.

V. SLOPE DISCONTINUITIES

We now derive the equations that govern the propagation of slope discontinuities. Let us consider again the weak (generalized) solutions to Eqs. (2) and (3), i.e., continuous solutions that may have discontinuities in their first derivatives in space. For simplicity, we first discuss Eq. (3). As in the previous section, the results derived from Eq. (3) (where $-\pi/2<\theta\leqslant\pi/2$) can be easily extended to the general case $(-\pi<\theta\leqslant\pi)$ of Eq. (2).

Suppose $(X(t),Y(t))$ describes the trajectory of a propagating slope discontinuity on the boundary curve $y=u(x,t)$, i.e., $Y(t)=u(X(t),t)$ and $u_x(X(t)-0,t)\neq u_x(X(t)+0,t)$. Here we use the shorthand notations

$$u_x(X(t)\pm 0,t)=\lim_{\epsilon\to 0} u_x(X(t)\pm\epsilon,t)$$

for the values of u_x (or other quantities) to the left $[x<X(t)]$ and right $[x>X(t)]$ of the slope discontinuity. Then, from Eq. (3), we have

$$u_t(X(t)-0,t)+f(t,X(t),Y(t),p_l)=0, \tag{17}$$

$$u_t(X(t)+0,t)+f(t,X(t),Y(t),p_r)=0, \tag{18}$$

where $f=c\sqrt{1+u_x^2}$ is the flux function defined in Eq. (4), and $p_l=u_x(X(t)-0,t)$ and $p_r=u_x(X(t)+0,t)$. Since the curve $y=u(x,t)$ is continuous at $x=X(t)$, we have the relation $u(X(t)-0,t)=u(X(t)+0,t)$. Thus, by differentiating this equation with respect to time t, we have

$$p_l\frac{dX(t)}{dt}+u_t(X(t)-0,t)=p_r\frac{dX(t)}{dt}+u_t(X(t)+0,t).$$

Substituting Eqs. (17) and (18) into the above equation, we obtain

$$\frac{dX(t)}{dt}=\frac{f_l-f_r}{p_l-p_r}, \tag{19}$$

where $f_l=f(t,X(t),Y(t),p_l)$ and $f_r=f(t,X(t),Y(t),p_r)$. Equation (19) represents the x component of the velocity of the slope discontinuity and is called the jump condition; in gas dynamics it is known as the Rankine–Hugoniot condition.

The y component of the velocity can also be easily determined. By differentiating $Y(t)=u(X(t)-0,t)$ $[=u(X(t),0)]$ with respect to time t, we obtain $dY(t)/dt=p_l\,dX(t)/dt+u_t(X(t)-0,t)$, i.e.,

$$\frac{dY(t)}{dt}=p_l\frac{dX}{dt}-f_l=p_r\frac{dX}{dt}-f_r=\frac{p_rf_l-p_lf_r}{p_l-p_r}, \tag{20}$$

where we have made use of Eqs. (17) and (19).

It is again straightforward to extend Eqs. (19) and (20) to the general case based on Eq. (2). Using the extended slope angle θ $(-\pi<\theta\leqslant\pi)$ defined in Eq. (15) and the relation $f=c/\cos\theta$ [Eq. (16)], we may rewrite Eqs. (19) and (20) as

$$\frac{dX(t)}{dt}=\frac{\cos\theta_lc_l-\cos\theta_rc_r}{\sin(\theta_l-\theta_r)}, \tag{21}$$

$$\frac{dY(t)}{dt}=\frac{\sin\theta_lc_l-\sin\theta_rc_r}{\sin(\theta_l-\theta_r)}, \tag{22}$$

where $c_l=c(t,x,y;p_l)$ and $c_r=c(t,x,y;p_r)$.

VI. THE ENTROPY CONDITION

The weak (generalized) solutions that we introduced in Sec. IV allow us to examine global behavior (i.e., for all time) of the "solutions" to Eqs. (2) and (3). In other words, even if the function $\phi(x,y,t)$ ceases to be differentiable at certain points after some time, it may still be considered a solution as long as it satisfies Eq. (2) "almost everywhere." As already noted, however, there are many generalized solutions for given initial conditions. The *entropy condition* is the criterion that identifies the solution with physical significance (the name comes from gas dynamics, in which the correct discontinuous solution across a shock is determined by the requirement that the entropy of the gas must increase).

We discuss the special case of Eq. (3), where the boundary surface is described by the function $y=u(x,t)$, and for simplicity we assume that the rate function depends only on the slope, i.e., $c=c(p)$. In this section, we briefly sketch the entropy condition and its application to our problem (a concise expository discussion on the entropy condition for the boundary evolution problem can be found in Ref. 3; for a more detailed mathematical discussion see, for example, Refs. 11 and 15).

The entropy condition for Eqs. (3) and (4) may be given as follows.[16] Suppose that a continuous function $u(x,t)$ satisfies Eq. (3) (except at the points where u_x is discontinuous). Then we choose this function $u(x,t)$ as the solution if and only if the first derivative $p=u_x$, in the neighborhood of each of its discontinuities, satisfies the following conditions.

If $p_r<p_l$, then the graph of $f(p) = c(p)\sqrt{1+p^2}$ over the interval $[p_r,p_l]$ in the $[p,f(p)]$ plane must lie *below* the chord connecting (p_r,f_r) and (p_l,f_l), i.e.,

$$f[\alpha p_r+(1-\alpha)p_l]<\alpha f_r+(1-\alpha)f_l, \tag{23}$$

for $0<\alpha<1$ [here $f_r=f(p_r)$ and $f_l=f(p_l)$ as before]. If $p_r>p_l$, on the other hand, the graph of $f(p)$ over $[p_r,p_l]$ must lie *above* the chord connecting (p_l,f_l) and (p_r,f_r), i.e.,

$$f[\alpha p_r+(1-\alpha)p_l]>\alpha f_r+(1-\alpha)f_l \tag{24}$$

for $0<\alpha<1$.

The two conditions (23) and (24) may be written as a single system of inequalities for all p between p_l and p_r:

$$\frac{f_r-f}{p_r-p}\lessgtr\frac{f_l-f_r}{p_l-p_r}\lessgtr\frac{f_l-f}{p_l-p}. \tag{25}$$

A slope discontinuity satisfying the entropy condition is called a facet corner—or a "shock" by analogy with shock waves in gas dynamics.

We now illustrate how the entropy condition is applied to the moving boundary problem. Suppose that, at time $t=0$, there is a slope discontinuity of $y=u(x,0)$, at which the slope $p=u_x$ changes from p_l to p_r ($p_r>p_l$) as x increases, as shown in Fig. 1. If the flux function $f(p) = c\sqrt{1+p^2}$ is convex, as in Fig. 2(a), the graph of $f(p)$ lies below the chord connecting the points (p_l,f_l) and (p_r,f_r), and condition (24) is violated. Therefore, no facet corner can develop from this slope discontinuity—instead, multiple characteristic lines emanate from the corner point A and the initial sharp corner evolves into a smooth arc. Such characteristics may be calculated from Eqs. (8)–(10) for all p in the range $p_l\leqslant p\leqslant p_r$, as shown in Fig. 2(b). Using the gas dynamics analogy again, we call a smooth curve developing from a sharp corner a "rarefaction wave."

On the other hand, if the flux function $f(p)$ is concave, as in Fig. 3(a), the graph of $f(p)$ lies above the chord between the points (p_l,f_l) and (p_r,f_r), and condition (24) is satisfied. Therefore, the function $u(x,t)$ with this slope discontinuity is the legitimate solution and the discontinuity (facet corner) propagates according to Eqs. (19) and (20). Figure 3(b) illustrates the characteristics and the propagation of the facet corner.

A more complicated example is shown in Fig. 4(a), where the flux function $f(p)$ is neither convex nor concave. In this case, we construct a convex envelope (concave envelope for the case $p_r<p_l$), as illustrated in Fig. 4(a). It is clear from this figure that two facet corners (shocks), associated with the intervals $[p_l,p^{(1)}]$ and $[p^{(2)},p_r]$ develop from the original single corner. Denoting the trajectories of these two facet corners by (X_1,Y_1) and (X_2,Y_2), we have, from Eqs. (19) and (20),

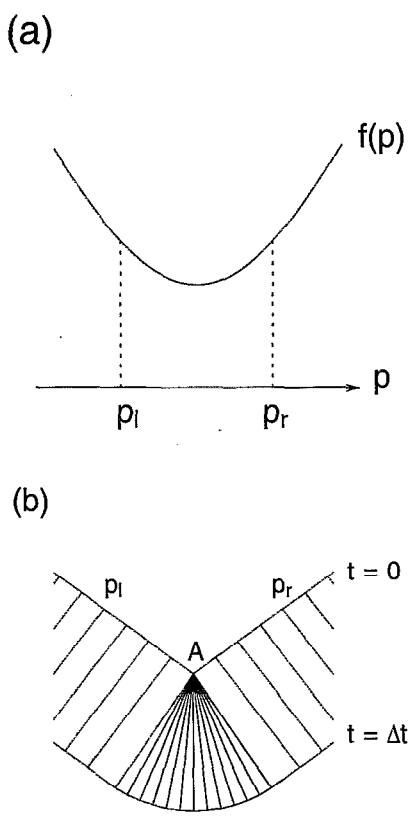

FIG. 2. (a) A convex flux function. (b) The characteristics and boundary curves at $t=0$ and $t=\Delta t>0$. Here $c_x=c_u=0$ are assumed.

$$\frac{dX_1(t)}{dt}=\frac{f^{(1)}-f_l}{p^{(1)}-p_l},\quad \frac{dY_1(t)}{dt}=\frac{p^{(1)}f_l-p_lf^{(1)}}{p^{(1)}-p_l},$$

and

$$\frac{dX_2(t)}{dt}=\frac{f_r-f^{(2)}}{p_r-p^{(2)}},\quad \frac{dY_2(t)}{dt}=\frac{p^{(2)}f_r-p_rf^{(2)}}{p_r-p^{(2)}},$$

where $f^{(1)}=f(p^{(1)})$ and $f^{(2)}=f(p^{(2)})$. Values of p between $p^{(1)}$ and $p^{(2)}$ give rise to a rarefaction wave propagating from the sharp corner of the initial curve, as in the case of Fig. 2. Figure 4(b) shows the resulting evolution of the surface in this case.

VII. THE NUMERICAL ALGORITHM

We now describe an algorithm that generates numerical solutions to the surface evolution problem, based on the theory of the Hamilton–Jacobi equation discussed above. We approximate the exact, piecewise-smooth boundary curve in a piecewise-linear manner, as illustrated in Fig. 5. The slope between consecutive nodes (x_i,y_i) and (x_{i+1},y_{i+1}) is denoted by $p_{i+1/2}$. In the calculations, it is more convenient to use the extended angle variable θ, rather than the slope p, since the former variable is bounded ($-\pi<\theta\leqslant\pi$) and can systematically handle arbitrary geometries, as discussed previously. From the definition (15), we introduce the discretized extended angle variable as follows:

$$\tan\theta_{i+1/2}=p_{i+1/2},$$

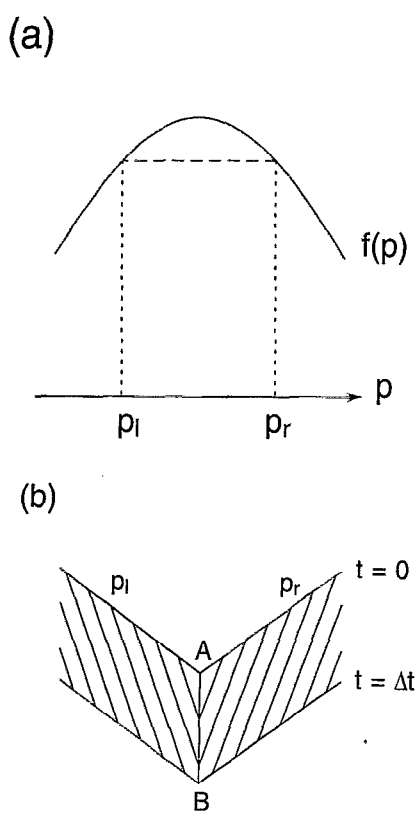

FIG. 3. (a) A concave flux function. (b) The characteristics and boundary curves at $t=0$ and $t=\Delta t>0$. The line AB represents the shock. Here $c_x=c_u=0$ are assumed.

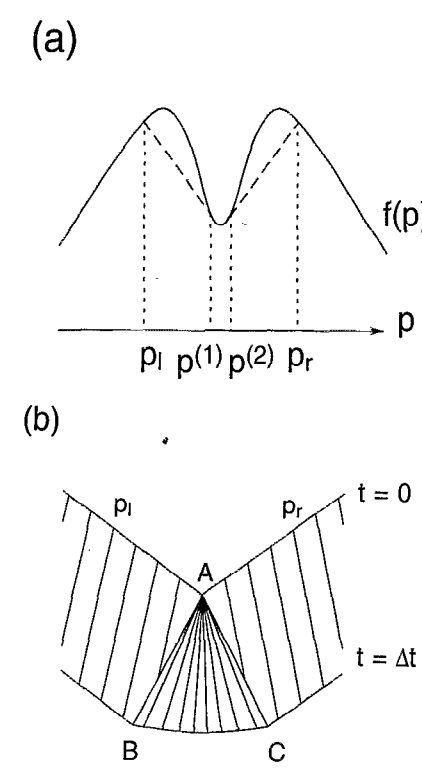

FIG. 4. (a) A flux function that is neither convex nor concave. (b) The characteristics and boundary curve at $t=0$ and $t=\Delta t>0$. The lines AB and AC represent the shocks. Here $c_x=c_u=0$ are assumed.

with

$$-\pi/2\leqslant\theta_{i+1/2}\leqslant\pi/2, \quad \text{if } x_i\leqslant x_{i+1},$$

$$-\pi<\theta_{i+1/2}\leqslant-\pi/2 \quad \text{or} \quad \pi/2\leqslant\theta_{i+1/2}\leqslant\pi, \quad \text{if } x_i\geqslant x_{i+1}.$$

In other words, $\theta_{i+1/2}$ is the angle that the line segment from (x_i,y_i) to (x_{i+1},y_{i+1}) makes with the positive x direction.

The central idea of our algorithm is to treat *every* node (x_i,y_i) of the piecewise-linear discretized curve as a facet corner, and to then calculate the shocks and rarefaction waves emanating from each node using the jump conditions (21) and (22) and the entropy condition (25). It will be shown that the propagation of discontinuities thus generated is also a good approximation to the propagation of the characteristics if the nodal point, together with the adjacent linear segments, approximate a neighborhood of a smooth point of the original curve.

A. Shocks

If only a single shock (facet corner) is known to emanate from the node (x_i^0,y_i^0) at time $t=0$, then one needs to use the jump conditions (21) and (22), instead of the characteristic equations, to evolve this facet corner. From Eqs. (21) and (22), the new position $(x_i^{\Delta t},y_i^{\Delta t})$ of the node at time $t=\Delta t$ is given by

$$x_i^{\Delta t}=x_i^0+\Delta t\frac{\cos\theta_l c_l-\cos\theta_r c_r}{\sin(\theta_l-\theta_r)}, \tag{26}$$

$$y_i^{\Delta t}=y_i^0+\Delta t\frac{\sin\theta_l c_l-\sin\theta_r c_r}{\sin(\theta_l-\theta_r)}, \tag{27}$$

where $\theta_l=\theta_{i-1/2}$, $\theta_r=\theta_{i+1/2}$, $c_l=c(t=0,x_i^0,y_i^0,\theta_{i-1/2})$ and $c_r=c(t=0,x_i^0,y_i^0,\theta_{i+1/2})$. Note that the rate function c used on the right-hand sides is evaluated on the initial point (x_i^0,y_i^0) at $t=0$. Thus, the integration scheme employed here is first order in time (higher-order schemes will be addressed in a future paper).

When the rate function c depends only on the slope θ, the slope of the segment between adjacent nodes (x_i,y_i) and (x_{i+1},y_{i+1}) will remain constant during the motion, i.e.,

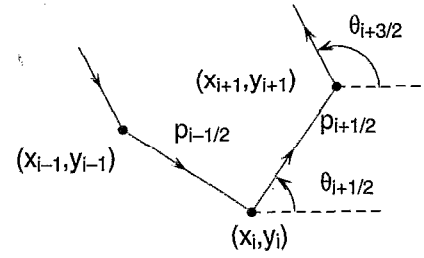

FIG. 5. Discretization of a curve by a piecewise-linear function. Here $\theta_{i+1/2}$ $(-\pi<\theta_{i+1/2}\leqslant\pi)$ is the angle that the line segment from (x_i,y_i) to (x_{i+1},y_{i+1}) makes with the positive x direction, and related to the slope by $p_{i+1/2}=\tan(\theta_{i+1/2})$.

$$\frac{y_{i+1}^{\Delta t}-y_i^{\Delta t}}{x_{i+1}^{\Delta t}-x_i^{\Delta t}}=\frac{y_{i+1}^0-y_i^0}{x_{i+1}^0-x_i^0}. \tag{28}$$

It is easily confirmed that Eq. (28) follows directly from Eqs. (26) and (27). Therefore, as in the case of characteristics, the trajectories of nodal points (x_i,y_i) are straight lines as long as they do not intersect other trajectories.

If the angles on each side of node (x_i,y_i) are equal, $\theta_{i-1/2}=\theta_{i+1/2}=\theta$, then the characteristic equations (11) and (12), instead of Eqs. (26) and (27), must be used:

$$x_i^{\Delta t}=x_i^0+\Delta t[c_\theta(\theta)\cos\theta+c(\theta)\sin\theta], \tag{29}$$

$$y_i^{\Delta t}=y_i^0+\Delta t[c_\theta(\theta)\sin\theta-c(\theta)\cos\theta]. \tag{30}$$

Again, the functions $c(\theta)=c(t,x,y,\theta)$ and $c_\theta=c_\theta(t,x,y,\theta)$ on the left-hand sides of Eqs. (29) and (30) are all evaluated with the initial data (x_i^0,y_i^0) and $t=0$. Note that the third equation (13) for the evolution of the angle θ is not solved here—instead, the extended angle variable θ is calculated at each time step from

$$\tan\theta_{i+1/2}=\frac{y_{i+1}-y_i}{x_{i+1}-x_i}$$

using the usual conventions.

While some nodal points actually represent true facet corners of the moving boundary, many of them are actually just a consequence of using piecewise-linear approximations to smooth segments [as the spatial discretization about (x_i,y_i) is continuously refined, this node represents a smooth point or a true corner according to whether or not $|\theta_{i+1/2}-\theta_{i-1/2}|\to 0$]. The evolution of segments that remain smooth over a given time interval should really be obtained from the characteristic equations (8)–(10), rather than the jump conditions (21) and (22). However, as we shall demonstrate below, the discretized shock equations (26) and (27) are also good approximations to the characteristic equations if the piecewise-linear approximation is in reasonably accurate agreement with the smooth curve.

Let θ be the extended angle variable of the exact smooth boundary at (x,y). Then, if a piecewise-linear locus with node $(x_i,y_i)=(x,y)$ approximates this smooth boundary, we may write $\theta_{i+1/2}=\theta+\delta_+$ and $\theta_{i-1/2}=\theta+\delta_-$, with $\delta_+\neq\delta_-$ and $|\delta_\pm|\ll 1$. From Eqs. (21) and (22), we obtain

$$\frac{dx_i}{dt}=\frac{\cos(\theta+\delta_+)c(\theta+\delta_-)-\cos(\theta+\delta_-)c(\theta-\delta_+)}{\sin(\delta_--\delta_+)}$$

$$=c_\theta(\theta)\cos\theta+c(\theta)\sin\theta+\mathcal{O}(|\delta_+-\delta_-|),$$

$$\frac{dy_i}{dt}=\frac{\sin(\theta+\delta_+)c(\theta+\delta_-)-\sin(\theta+\delta_-)c(\theta-\delta_+)}{\sin(\delta_--\delta_+)}$$

$$=c_\theta(\theta)\sin\theta-c(\theta)\cos\theta+\mathcal{O}(|\delta_+-\delta_-|).$$

Thus, to first order in the angle discretization, the equations of motion for (x_i,y_i) regarded as a shock (i.e., a point of slope discontinuity) agree with the characteristic equations (11) and (12) for the motion of (x_i,y_i) regarded as a smooth boundary point. The same can be seen by comparing Eq. (19) with Eq. (8), and Eq. (20) with Eq. (9), in the limit $p_l\to p_r$.

When two shock trajectories intersect within a single time step Δt, we use the following procedure. Let the rate function c depend only on the angle θ. Since the slope remains constant and the shock trajectories are straight lines during the motion, the time interval Δt^* after which two shocks intersect is easily determined. By setting $x_{i+1}^{\Delta t}=x_i^{\Delta t}$ and solving this for Δt, we obtain

$$\Delta t^*=-\Delta x/\Delta u, \tag{31}$$

where $\Delta x=x_{i+1}-x_i$ and $\Delta u=u_{i+1}-u_i$, and

$$u_i=\frac{\cos\theta_{i+1/2}c(\theta_{i-1/2})-\cos\theta_{i-1/2}c(\theta_{i+1/2})}{\sin(\theta_{i-1/2}-\theta_{i+1/2})}. \tag{32}$$

Of course, the same expression for Δt^* could also be obtained from the equation $y_{i+1}^{\Delta t}=y_i^{\Delta t}$. If $0<\Delta t^*<\Delta t$, the time step Δt^* is used instead of Δt, and the intersection point $(x_i^{\Delta t^*},y_i^{\Delta t^*})$ becomes a new node that replaces the former nodes (x_i,y_i) and (x_{i+1},y_{i+1}).

If the rate function c depends on the position (x,y) and time t, the slope between adjacent nodes (x_i,y_i) and (x_{i+1},y_{i+1}) is no longer constant, i.e., Eq. (28) does not hold. Consequently, the true shock trajectories are no longer straight lines and the intersection time Δt^* cannot be defined by Eq. (31) in the strict sense. However, by taking the time step Δt sufficiently small, one can force Eq. (28) to hold approximately, but with high accuracy. Within this small time step, one may apply the same method as the one presented above to calculate intersection of shocks for the case of $c=c(t,x,y,\theta)$.

B. Rarefaction waves

Suppose that the entropy condition indicates that a single rarefaction wave develops from the node (x_i^0,y_i^0) given at time $t=0$, as in the example of Fig. 2. (The numerical implementation of the entropy condition will be discussed in the next section.) Then, using a predetermined angle resolution $\Delta\theta$, we choose a set of angle variables $\{\theta^{(n)}\}$ $(0\leqslant n\leqslant N)$ such that $\theta_l=\theta^{(0)}<\theta^{(1)}<\theta^{(2)}<\cdots<\theta^{(N)}=\theta_r$ with $\theta^{(n+1)}-\theta^{(n)}\simeq\Delta\theta$, where $\theta_l<\theta_r$ is assumed ($\theta_l=\theta_{i-1/2}$ and $\theta_r=\theta_{i+1/2}$). With these discretized angles, the node (x_i,y_i) may be regarded as the infinitesimal limit of a "round" corner, consisting of the set of angles $\{\theta^{(n)}\}$. For each of such angles $\{\theta^{(n)}\}$, we solve the shock-trajectory equations (21) and (22): the position of the new node $(x_{i,n+1/2}^{\Delta t},y_{i,n+1/2}^{\Delta t})$ associated with two adjacent angles $\theta^{(n)}$ and $\theta^{(n+1)}$ after a time step Δt is given by

$$x_{i,n+1/2}^{\Delta t}=x_i^0+\Delta t\frac{\cos\theta^{(n+1)}c(\theta^{(n)})-\cos\theta^{(n)}c(\theta^{(n+1)})}{\sin(\theta^{(n)}-\theta^{(n+1)})}, \tag{33}$$

$$y_{i,n+1/2}^{\Delta t}=y_i^0+\Delta t\frac{\sin\theta^{(n+1)}c(\theta^{(n)})-\sin\theta^{(n)}c(\theta^{(n+1)})}{\sin(\theta^{(n)}-\theta^{(n+1)})}. \tag{34}$$

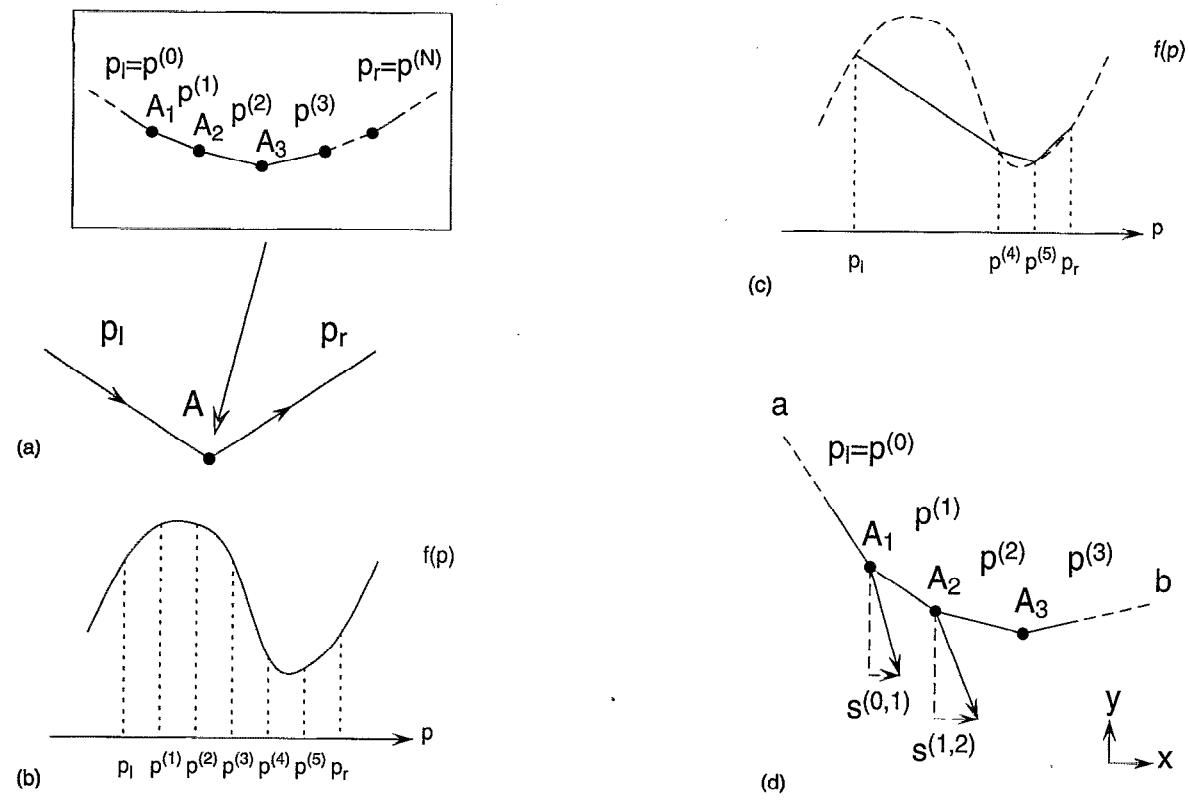

FIG. 6. (a) A facet corner of the curve, which may be viewed as the limit of a round corner (inset). (b) The flux function $f(p)$ with discrete values of p. (c) The convex envelope constructed from the discrete values of p. (d) The set of line segments $aA_1A_2A_3b$ represents the boundary curve shown in the inset of (a). The arrows from A_1 and A_2 are the shocks, whose horizontal components are given by $s^{(0,1)}$ and $s^{(1,2)}$.

The functions $c(\theta)=c(t,x,y,\theta)$ on the right-hand sides are evaluated at the initial data (x_i^0,y_i^0) and $t=0$, as before.

C. The entropy condition

We now discuss the numerical implementation of the entropy condition (25). For simplicity, we consider here only the system defined by Eq. (3) with the flux function (i.e., sputtering yield) $f(p)=c(p)\sqrt{1+p^2}$. An algorithm implementing the entropy condition for the general case of Eq. (2) may be found in the Appendix.

Suppose the node (x_i,y_i) has neighboring slopes $p_l=\tan\theta_l$ and $p_r=\tan\theta_r$ with $p_l<p_r$, as shown in Fig. 6(a). We are interested in determining how this node will develop into shocks and/or rarefaction waves, based on the entropy condition (25). As in the case of the rarefaction waves, we introduce a set of angle variables $\{\theta^{(n)}\}$ $(0\leqslant n\leqslant N)$, such that $\theta_l=\theta^{(0)}<\theta^{(1)}<\theta^{(2)}<\cdots<\theta^{(N)}=\theta_r$, using the predetermined angle resolution $\Delta\theta$ $(\theta^{(n+1)}-\theta^{(n)}\simeq\Delta\theta)$. With $p^{(n)}=\tan\theta^{(n)}$, the flux function $f(p)$ may be discretized, as in Fig. 6(b). Since $p_l<p_r$, we need to construct the convex envelope of the flux function over $[p_l,p_r]$, as shown in Fig. 6(c). The method to construct such an envelope is as follows.

(1) First calculate the slopes of the flux function $f(p)$ over $[p^{(0)},p^{(1)}]$ and $[p^{(1)},p^{(2)}]$, i.e.,

$$s^{(i,j)}=\frac{f(p^{(i)})-f(p^{(j)})}{p^{(i)}-p^{(j)}}, \tag{35}$$

with $(i,j)=(0,1)$ and $(1,2)$.

(2) If the slope over $[p^{(0)},p^{(1)}]$ is less than that over $[p^{(1)},p^{(2)}]$ (i.e., $s^{(0,1)}<s^{(1,2)}$), then we decide to keep the point $[p^{(1)},f(p^{(1)})]$, at least temporarily, and proceed to the next step, i.e., compare the slope $s^{(1,2)}$ over $[p^{(1)},p^{(2)}]$ with the next adjacent slope $s^{(2,3)}$ over $[p^{(2)},p^{(3)}]$.

(3) If the slope over $[p^{(0)},p^{(1)}]$ is greater than or equal to that over $[p^{(1)},p^{(2)}]$ (i.e., $s^{(0,1)}\geqslant s^{(1,2)}$), then we discard the point $[p^{(1)},f(p^{(1)})]$ and draw a straight line connecting $[p^{(0)},f(p^{(0)})]$ and $[p^{(2)},f(p^{(2)})]$. In the next step, we compare the slope $s^{(0,2)}$ over $[p^{(0)},p^{(2)}]$ with the next adjacent slope $s^{(2,3)}$ over $[p^{(2)},p^{(3)}]$.

(4) In general, we compare the slope $s^{(i,j)}$ over $[p^{(i)},p^{(j)}]$ $(i<j)$ with its right adjacent slope $s^{(j,j+1)}$ over $[p^{(j)},p^{(j+1)}]$. If $s^{(i,j)}<s^{(j,j+1)}$, then we proceed to the next step; compare $s^{(j,j+1)}$ with $s^{(j+1,j+2)}$. If $s^{(i,j)}\geqslant s^{(j,j+1)}$, then we discard the point $[p^{(j)},f(p^{(j)})]$ and create a new slope $s^{(i,j+1)}$ over $[p^{(i)},p^{(j+1)}]$. Before proceeding to the next step, we need to check if the new slope $s^{(i,j+1)}$ is greater than its left adjacent slope $s^{(m,i)}$ over $[p^{(m)},p^{(i)}]$. If not, then we also discard the point $[p^{(i)},f(p^{(i)})]$ and create a new step $s^{(m,j+1)}$. Repeat this process until either the new slope becomes greater than its left adjacent slope or the left end point of the new slope reaches the left end point $[p_l,f(p_l)]$ of the interval. Then we proceed to the next step; compare the new slope with its right adjacent slope $s^{(j+1,j+2)}$.

(5) These processes are repeated until the right end point $[p_r,f(p_r)]$ of the interval is reached. In the case of

$p_l > p_r$, we need to construct the concave envelope, instead of convex envelope, and the above algorithm ought to be modified accordingly. A more rigorous description of this algorithm with use of the angle variable θ, rather than p, will be presented in the Appendix.

In the case of the flux function in Fig. 6(b), the approximated convex envelope is given in Fig. 6(c). Corresponding to the three chords over $[p_l,p^{(4)}]$, $[p^{(4)},p^{(5)}]$ and $[p^{(5)},p_r]$, therefore, we conclude that three shocks emanate from the node (x_i,y_i), based on Eqs. (26) and (27). (In this particular example, however, the shocks corresponding to the chords over $[p^{(4)},p^{(5)}]$ and $[p^{(5)},p_r]$ may be interpreted as approximations of rarefaction waves, rather than shocks.)

In the case of convex flux functions, such as the one in Fig. 2(a), no discretized points $[p^{(n)},f(p^{(n)})]$ between p_l and p_r are removed by the entropy condition since $s^{(n,n+1)} > s^{(n+1,n+2)}$, i.e., the slope over $[p^{(n)},p^{(n+1)}]$ is always less than that over $[p^{(n+1)},p^{(n+2)}]$. The rarefaction wave is thus obtained, as discussed in the previous section. Evidently, by taking the angular grid size $\Delta\theta$ sufficiently small, one can approximate flux functions $f(p)$ and their envelopes with sufficient accuracy.

It is instructive to examine the geometrical meaning of the entropy condition with the discretized slopes $p^{(n)}$. As in the case of the rarefaction waves discussed above, we regard the node $A=(x_i,y_i)$ as the infinitesimal limit of a "round" corner consisting of points $A^{(n)}$, $n=1,2,3,...$ [see the inset of Fig. 6(a)]. Let $p^{(n)}$ represent the slope between two points $A^{(n)}$ and $A^{(n+1)}$. The shock trajectory from each point $A^{(n)}$ can be calculated from Eqs. (19) and (20). Equation (19) indicates that the quantities $s^{(0,1)}$ and $s^{(1,2)}$ (i.e., the slopes over the intervals $[p^{(0)},p^{(1)}]$ and $[p^{(1)},p^{(2)}]$ in the p-$f(p)$ plane) represent the x components of the velocities of points $A^{(1)}$ and $A^{(2)}$. Since the distance between these two points $A^{(1)}$ and $A^{(2)}$ is infinitesimally small, it is clear that, if $s^{(0,1)} > s^{(1,2)}$, then the two shock trajectories intersect immediately [Fig. 6(d)]: a new shock trajectory must be calculated from the two slopes $p^{(0)}(=p_l)$ and $p^{(1)}$ and compared with the right adjacent shock emanating from point $A^{(3)}$. On the other hand, if $s^{(0,1)} < s^{(1,2)}$, the two trajectories from $A^{(1)}$ and $A^{(2)}$ do not intersect each other, and we now need to check if the trajectory from point $A^{(2)}$ intersects the trajectory from point $A^{(3)}$. It is easy to confirm that this geometrical process to find the "surviving" shock trajectories from node A is equivalent to the entropy-condition algorithm presented above. For the case of the flux function given in Fig. 6(b), three trajectories are found to "survive," as discussed before.

For a description of the entropy-condition algorithm appropriate to the general surface-evolution equation (2), see the Appendix.

D. Geometric interpretation

We now briefly discuss the geometric meaning of the shock trajectories obtained from Eqs. (26) and (27). For simplicity, we again consider a rate function depending only on the slope, i.e., $c=c(\theta)$. Figure 7 shows the motion

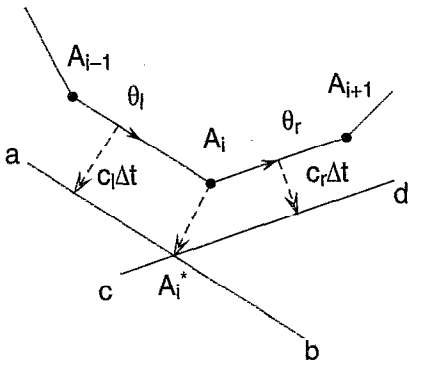

FIG. 7. The nodes $A_{i-1}=(x_{i-1}^0,y_{i-1}^0)$, $A_i=(x_i^0,y_i^0)$,..., represent the boundary curve at $t=0$. Here $\theta_l=\theta_{i-1/2}$ and $\theta_r=\theta_{i+1/2}$. The lines ab and cd are parallel to the line segments $A_{i-1}A_i$ and A_iA_{i+1} with distances $c_l\,\Delta t$ and $c_r\,\Delta t$, respectively.

of the line segments $A_{i-1}A_i$ and A_iA_{i+1}. Since the perpendicular velocity c of the line segments depends only on the slope, the line segments remain parallel to their original orientations: after a time step Δt, the segment $A_{i-1}A_i$ must lie somewhere on the line ab, and the segment A_iA_{i+1} somewhere on the line cd. It is easy to show that the new node $(x_i^{\Delta t},y_i^{\Delta t})$ calculated from Eqs. (26) and (27) is given by the intersection A_i^* of ab and cd. Namely, the vector $A_iA_i^*$ represents a possible shock trajectory that emanates from the node $A_i=(x_i^0,y_i^0)$. As we have discussed above, however, the vector $A_iA_i^*$ is not necessarily the correct shock solution, since many other shocks associated with the angular range between θ_l and θ_r also need to be taken into account, based on the entropy condition. If the difference between the two adjacent angles θ_l and θ_r is smaller than the angular grid size $\Delta\theta$, then the intersection A_i^* will be adopted as the new node corresponding to A_i.

VIII. REPRESENTATIVE EXAMPLES

Using the algorithm described above, we now present some representative examples of numerical calculations. In most examples shown in this section, the slope dependence of the rate function (i.e., etch rate) c is given by $c_0(\theta)$, as defined by

$$f_0(\theta)=\frac{c_0(\theta)}{\cos\theta}=1.454\,743\cos\theta-0.464\,719\cos 3\theta+0.015\,573\cos 5\theta-0.005\,669\cos 7\theta-0.010\,000\cos 9\theta+0.010\,552\cos 11\theta-0.006\,204\cos 13\theta+0.005\,725\cos 15\theta,$$

if $-\pi/2<\theta<\pi/2$, and $f_0(\theta)=c_0(\theta)=0$ otherwise. Note that $f_0(0)=1$. The shape of this function is similar to the one shown in Fig. 4(a). The flux function $f_0(\theta)$ is proportional to that used in Ref. 3.

Figure 8 shows the inert ion beam etching of a trench whose cross section is a semicircle at time $t=0$. The beam is assumed to be vertical and the etch rate $c=c_0(\theta)$ is uniform in space. The masks are assumed to be nonerod-ing. Some selected characteristics (i.e., shocks from nodal points of the discretized curve) are also shown in the do-

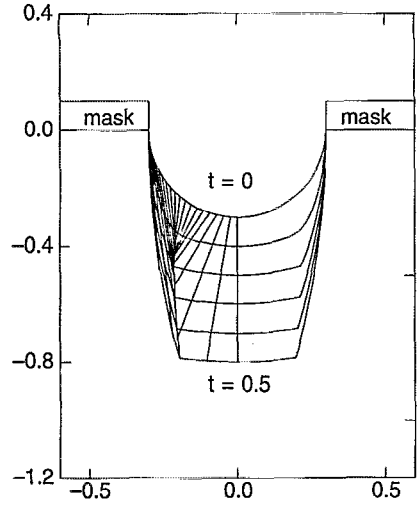

FIG. 8. The evolution of a semicircular trench due to inert ion beam etching. The etch rate $c=c_0(\theta)$, where $c_0(\theta)$ is given in Sec. VIII. The boundary profiles are shown at different times ($0<t<0.5$) corresponding to a constant time interval $\Delta t=0.1$. The masks are assumed to be noneroding. Some selected characteristics are shown in the domain $x<0$.

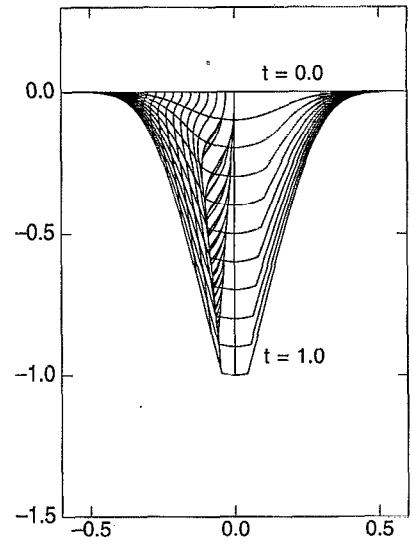

FIG. 10. The evolution of an initially flat surface due to inhomogeneous beam etching. The etch rate used here is $c(x,\theta)=c_0(\theta)\exp-(x/\sigma)^2$, with $\sigma=0.2$. Some selected characteristics are shown in the domain $x<0$.

main $x<0$ to illustrate the formation of facet corners; the configuration is symmetric about $x=0$. More characteristics (i.e., shocks) than those actually shown in Fig. 8 were used to obtain the boundary profiles in this calculation. The boundary profiles are shown at equal time intervals $\Delta t=0.1$ for $0<t<0.5$. Evidently, two shocks (i.e., facet corners) are formed as a result of the merging of characteristics.

Figure 9 shows the evolution of a rectangular trench with the same rate function $c=c_0(\theta)$. Some selected characteristics are shown for $x<0$. No masks are present in this case, and the facet corners denoted by α, β and λ, κ develop as a result. Rarefaction waves are also visible at the sections $\alpha-\beta$ and $\kappa-\lambda$. It is also seen that the bottom of the trench becomes narrower as it evolves, creating a tapered profile. This is due to the fact that the flux function $f_0(\theta)$ approaches 0 relatively fast as $\theta\to\pm\pi/2$. As opposed to tapering induced by sidewall passivation, we call this *intrinsic* tapering, i.e., tapering caused solely by the θ dependence of the etch rate function $c(\theta)$.

Figure 10 is an example of spatially inhomogeneous ion beam etching of a flat surface. The etch rate here is given by

$$c(x,\theta)=c_0(\theta)\exp(-x^2/\sigma^2),$$

with $\sigma=0.2$. Some selected characteristics are shown for $x<0$; the formation of two facet corners is clearly visible.

In Fig. 11 we show a model of purely chemical etching, with an etch rate that is spatially discontinuous. The etch

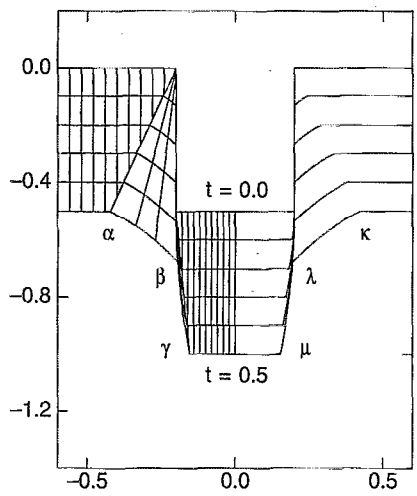

FIG. 9. The evolution of a rectangular trench without masks under the same conditions as those in Fig. 8. Some selected characteristics are shown in the domain $x<0$.

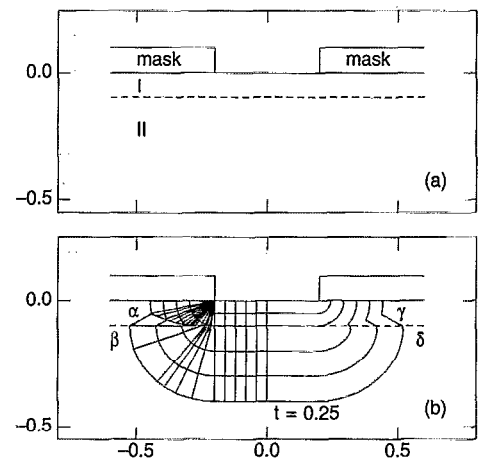

FIG. 11. Isotropic etching of layered materials. The etch rates are $c=1$ in the material I and $c=2$ in the material II. (a) The initial profile. (b) The evolution of the profile. Some selected characteristics are shown in the domain $x<0$.

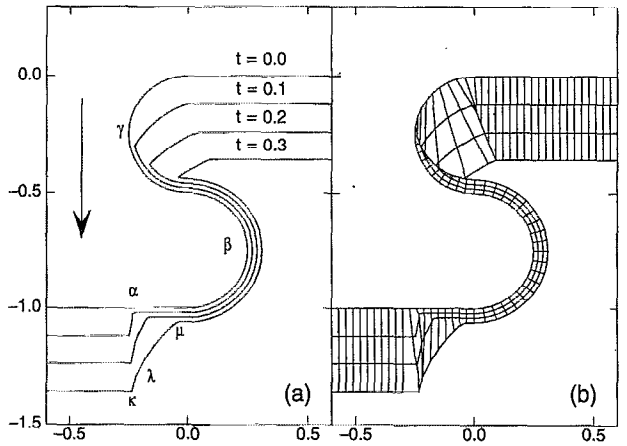

FIG. 12. (a) Reactive ion etching with a shadowing (i.e., visibility angle) effect. The etch rate is given by $c=c_0(\theta)+0.2$ if the surface is exposed to the incoming etchant beam, and $c=0.2$ if not. (b) Some selected characteristics.

rate here is assumed to be isotropic (i.e., independent of the slope), but dependent on position. Figure 11(a) shows the profile at $t=0$. There are two materials with different etch rates: $c=1$ for material I, and $c=2$ for material II. Noneroding masks partially cover material I. Figure 11(b) shows the evolution of the boundary profile at equal time intervals $\Delta t=0.05$, with selected characteristics shown for $x<0$. The formation of the facet corners α and γ is due to the interaction of two groups of characteristics, one of which represents a rarefaction wave emanating from the mask corners, while the other corresponds to characteristics originating from the boundary between materials I and II. Mask undercut due to isotropic etching is evident in Fig. 11(b). Note that no particular conditions are applied to the characteristics or shocks at the material interface between I and II in this example. Our numerical scheme automatically emanates necessary characteristics from the interface.

Figure 12 shows a model of reactive ion etching, where the etch rate is given as a sum of isotropic and nonisotropic components. The initial shape of the material is given by the curve denoted by $t=0$ in Fig. 12(a). As in the previous examples, the incoming etchant (i.e., reactive ion) beam is assumed to be vertical [as indicated by an arrow in Fig. 12(a)], and no reflection of the etchants or redeposition of sputtered materials are taken into account. Therefore, the part of the initial surface indicated by $\alpha-\beta-\gamma$ in Fig. 12(a) is not exposed to the directional etchant beam, and is subject only to isotropic chemical etching. In other words, the etch rate here is given by

$$c=c_0(\theta)+0.2,$$

if the surface is exposed to the beam, and $c=0.2$ otherwise. This shadowing (i.e., visibility angle) effect results in the development of the three facet corners κ, λ, and μ from the initial point α. The procedure for calculating visibility angles used in this calculation will be discussed elsewhere. Selected characteristics are shown in Fig. 12(b).

IX. DISCUSSION AND CONCLUSIONS

A new surface-evolution algorithm, based on the Hamilton–Jacobi equation (2) and entropy condition (25), has been presented. In this method, which may be characterized as a shock-tracking method, all shocks (i.e., facet corners) and rarefaction waves emanating from slope discontinuities are obtained directly from the entropy condition, and their propagation is computed from the jump (Rankine–Hugoniot) condition, i.e., Eqs. (21) and (22). The propagation of characteristics is shown to be well approximated by the jump condition.

Since we have used the lowest-order finite difference scheme in evaluating space gradients, the surface (curve) representation used in this method is the same as that of most "geometric" models—a sequence of nodal points joined by linear segments. This facilitates a comparison of the shock-tracking method with the more-naive geometric models. The shock-tracking method, which has been derived from first principles based on a well-defined physical system, indicates the "correct" way of advancing the nodes and segments that represent an evolving surface.

Unlike most characteristics calculations of evolving surfaces, the shock-tracking method requires no geometric adjustments, such as "delooping," when intersections of characteristics occur. The entropy condition implemented in the shock-tracking method automatically selects the physically meaningful solution. As mathematical theory of the Hamilton–Jacobi equation[17–19] and systems of conservation laws[11,15] indicates, the equation of motion of the surface, Eq. (2) or Eq. (3), has a unique "weak" solution if and only if the entropy condition (25) is satisfied. As noted in Refs. 1 and 2, calculating the locus of the characteristics or employing Huygens' principle does not provide a sufficient criterion for eliminating nonphysical solutions such as loops.

Numerical solutions of the initial-boundary problem for PDE (5) with a sufficiently small diffusion constant ν provide a good approximation to the viscosity solution.[5] Various numerical techniques to solve such problems have been developed, especially in the context of gas dynamics and systems of conservation laws. Typical finite-difference methods employ the combination of a Lax–Wendroff-type scheme and the upwind scheme to deal with large gradients of the solutions. Even if ν is set to zero in Eq. (5), such finite-difference schemes introduce a "numerical" diffusion, i.e., they effectively include a diffusion term on the right-hand side. While various schemes have been devised to minimize this numerical diffusion and treat the shocks with sufficient accuracy, the shock-tracking method discussed in this paper explicitly solves for the slope discontinuities without any diffusion effects.

Although the PDE approach based on Eq. (5) can easily accommodate the complicated geometry of a moving boundary, the equation needs to be solved over the *entire* x-y plane as an initial-boundary-value problem, rather than specifically on the location of the moving curve $\phi(t,x,y)=0$. This makes the PDE calculation less efficient in case of relatively simple geometries, such as the boundary curves observed in reactive ion etching.

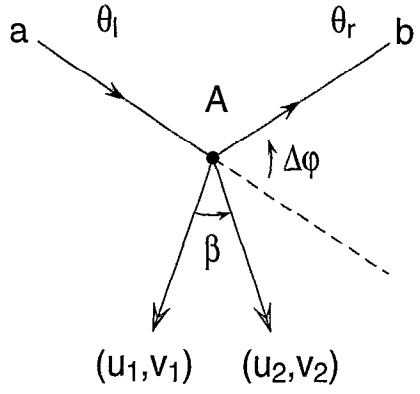

FIG. 13. The line segments aAb represent the boundary curve at $t=0$. The vectors (u_1,v_1) and (u_2,v_2) denote the velocities of shocks emanating from A.

Solving the PDE on a (fixed) spatial grid also requires the rate function c to be defined on the entire x-y plane. Although this was true in most of the simple examples given above, it is not necessarily the case when more realistic physical or chemical considerations are taken into account. For example, if the etch rate is influenced by the shadowing (i.e., visibility-angle effect) of the incoming ion flux, then c is defined only on the boundary curve $\phi=0$. As shown in Fig. 12, such a situation can be easily handled by the shock-tracking method. The problem of the rate function may be avoided by solving Eq. (6) [or (3)], as shown in Ref. 3, but the geometry of an evolving curve is then limited to forms describable by $y=u(x,t)$, which precludes the possibility of "undercuts."

The shock-tracking method is thus especially suited to the surface evolution problem in reactive ion etching, where sharp corners are known to develop, even from a smooth initial profile. By reversing the sign of the rate function c, the method can also be applied to certain deposition problems (e.g., a realistic simulation of reactive ion etching might have to incorporate a simultaneous redeposition flux of sputtered materials). These issues will be addressed in a future paper.

APPENDIX: THE ENTROPY CONDITION FOR θ

Here we extend the entopy-condition algorithm discussed in Sec. VII C to the general case of Eq. (2), using the extended slope angle variable θ ($-\pi<\theta\leqslant\pi$). Consider the node A, whose left and right adjacent angles variables are θ_l and θ_r ($\theta_l\neq\theta_r$), as shown in Fig. 13. Choose an angle θ_1 ($-\pi<\theta_1\leqslant\pi$) between the two angles θ_l and θ_r in the sense that $\theta_1\equiv\theta_l+\alpha\,\Delta\varphi(\mathrm{mod}\,2\pi)$. Here $0<\alpha<1$, and $\Delta\varphi\equiv\theta_r-\theta_l(\mathrm{mod}\,2\pi)$ ($-\pi<\Delta\varphi<\pi$) denotes the difference between the two angles of the adjacent line segments. The shock trajectories associated with the pairs of angles (θ_l,θ_1) and (θ_1,θ_r) may be calculated from Eqs. (21) and (22): their velocity vectors are given by (u_1,v_1) and (u_2,v_2), where

$$u_1=\frac{\cos\theta_1 c(\theta_l)-\cos\theta_l c(\theta_1)}{\sin(\theta_l-\theta_1)},$$

$$v_1=u_1\tan\theta_1-\frac{c(\theta_1)}{\cos\theta_1}=\frac{\sin\theta_1 c(\theta_l)-\sin\theta_l c(\theta_1)}{\sin(\theta_l-\theta_1)},$$

and similar formulas hold for (u_2,v_2). The angle β formed by these velocity vectors is given by

$$\sin\beta=\frac{u_1v_2-u_2v_1}{\sqrt{(u_1^2+v_1^2)(u_2^2+v_2^2)}}=\frac{(u_2-u_1)f(\theta_1)}{\sqrt{(u_1^2+v_1^2)(u_2^2+v_2^2)}},$$

where the relations $f(\theta)=c(\theta)/\cos\theta$ [Eq. (16)], $v_1=\tan\theta_1 u_1-f(\theta_1)$ and $v_2=\tan\theta_1 u_2-f(\theta_1)$ are used. If β is sufficiently small (or, more loosely, $-\pi/2<\beta<\pi/2$), the sign of β is equal to the sign of $(u_2-u_1)\cos\theta_1 c(\theta_1)$. (In our system, $c>0$ and $c<0$ indicate "etching" and "deposition," respectively.)

It is clear from the geometric discussion on shock trajectories emanating from a single point [Fig. 6(a)] that if the angle difference β is positive, then the two shock trajectories do not intersect. On the other hand, if β is negative, then the two shocks intersect each other at the moment they start to propagate, and a new shock trajectory must be constructed with the use of the slopes θ_l and θ_r.

It is straightforward to generalize this argument to the case where we have many discretized angles $\{\theta^{(n)}\}$ ($n=1,2,\ldots$) between θ_l and θ_r. If $\Delta\varphi\equiv\theta_r-\theta_l(\mathrm{mod}\,2\pi)$ defined above is positive, then the numerical implementation of the entropy condition may be given by the following algorithm.

(1) Choose the discretized angles $\{\theta^{(n)}\}$ ($n=1,2,\ldots,N$) based on the predetermined angular grid size $\Delta\theta$, i.e.,

$$\theta_l<\theta^{(1)}<\theta^{(0)}<\cdots<\theta^{(N)},$$

where $\theta_N\equiv\theta_r(\mathrm{mod}\,2\pi)$ and $\theta^{(n+1)}-\theta^{(n)}\simeq\Delta\theta$. Set $\theta^{(0)}=\theta_l$ and let $[\theta^{(i)},\theta^{(j)}]$ denote the interval of θ given by $\theta^{(i)}\leqslant\theta\leqslant\theta^{(j)}$.

(2) We now inductively define the set

$$\mathcal{S}_n=\{[\theta^{(0)},\theta^{(j_n(1))}],[\theta^{(j_n(1))},\theta^{(j_n(2))}],\ldots,$$
$$\times[\theta^{(j_n(m(n)-1))},\theta^{(j_n(m(n)))}]\}$$

of $m(n)$ intervals [$1\leqslant m(n)\leqslant N$] for each n ($1\leqslant n\leqslant N2$). Here $m(n)$ and $j_n(k)$ [$k=0,1,2,\ldots,m(n)$] denote integer functions of n and k, respectively. The function $j_n(k)$ also depends on n ($1\leqslant n\leqslant N$), satisfying the relation $j_n(m(n))=n$ for all n. The definition of $\mathcal{S}_n$ is given as follows:

(i) Set $\mathcal{S}_0=\{[\theta^{(0)},\theta^{(1)}]\}$.

(ii) Suppose that

$$\mathcal{S}_n=\{[\theta^{(0)},\theta^{(j_n(1))}],[\theta^{(j_n(1))},\theta^{(j_n(2))}],\ldots,$$
$$\times[\theta^{(j_n(m(n)-1))},\theta^{(j_n(m(n)))}]\}$$

is given, where $m(n)\geqslant 1$. Then define $\mathcal{S}_{n+1}$ from the following procedure.

(a) Consider the set of integers

$$\{0,j_n(1),j_n(2),\ldots,j_n(m(n)),n+1\}.$$

Let p be a positive integer and set $p:=m(n)$.

(b) Calculate

$$\mathscr{L}=(s^{(n+1,j_n(p))}-s^{(j_n(p),j_n(p-1))})\cos\theta^{j_n(p)}c(\theta^{j_n(p)}),$$

where $s^{(i,j)}$ is the same quantity defined in Eq. (35), i.e.,

$$s^{(i,j)}=\frac{\cos\theta^{(i)}c(\theta^{(j)})-\cos\theta^{(i)}c(\theta^{(j)})}{\sin(\theta^{(j)}-\theta^{(i)})}.$$

(c) If $\mathscr{L}>0$, then define $m(n+1)=p+1$ and a new set of integers,

$$\{0,j_{n+1}(1),j_{n+1}(2),\ldots,j_{n+1}(p),j_{n+1}(p+1)\}$$
$$:=\{0,j(1),j_n(2),\ldots,j_n(p),n+1\}. \quad \text{(A1)}$$

Note that $j_{(n+1)}(m(n+1))=j_{(n+1)}(p+1)=n+1$. This is the end of procedure in defining $\mathscr{S}_n$; proceed to (iii).

(d) If $\mathscr{L}<0$, then there are two possibilities.

(d1) If $p=1$, then define $m(n+1)=1$ and $j_{n+1}(m(n+1))=j_{n+1}(1)=n+1$. In other words, a new set of integers $\{0,j_{n+1}(k)\}$ with $k=1=m(n+1)$ becomes

$$\{0,n+1\}. \quad \text{(A2)}$$

This is the end of procedure in defining $\mathscr{S}_n$; proceed to (iii).

(d2) If $p\geqslant 2$, then set $p:=p-1$ and return to (b).

(iii) Using the newly defined set of integers $\{0,j_{n+1}(k)\}$ $[1\leqslant k\leqslant m(n+1)=p+1]$ given in (A1) or (A2), we define

$$\mathscr{S}_{n+1}=\{[\theta^{(0)},\theta^{(j_{n+1}(1))}],[\theta^{(j_{n+1}(1))},\theta^{(j_{n+1}(2))}],\ldots,$$
$$\times[\theta^{(j_{n+1}(m(n+1)-1))},\theta^{(j_{n+1}(m(n+1)))}]\}.$$

(3) From the above inductive definition for $\mathscr{S}_n$, we eventually obtain

$$\mathscr{S}_N=\{[\theta^{(0)},\theta^{(j_N(1))}],[\theta^{(j_N(1))},\theta^{(j_N(2))}],\ldots$$
$$\times[\theta^{(j_N(m(N)-1))},\theta^{(j_N(m(N)))}]\}.$$

The $m(N)$ shocks associated with the pairs of angles $(\theta^{(j_N(i))},\theta^{(j_N(i+1))})$ $[0\leqslant i\leqslant m(N)]$ with $\theta^{(j_N(0))}=\theta^{(0)}=\theta_l$ emanate from the nodal point.

[1] R. Smith, S. J. Wilde, G. Carter, I. V. Katardjiev, and M. J. Nobes, J. Vac. Sci. Technol. B **5**, 579 (1987).
[2] I. V. Katardjiev, G. Carter, and M. J. Nobes, J. Phys. D **22**, 1813 (1989).
[3] D. S. Ross, J. Electrochem. Soc. **135**, 1235 (1988).
[4] D. S. Ross, J. Electrochem. Soc. **135**, 1260 (1988).
[5] S. Osher and J. A. Sethian, J. Comput. Phys. **79**, 12 (1988).
[6] J. Glimm, Comments Pure Appl. Math. **18**, 697 (1965).
[7] A. J. Chorin, J. Comput. Phys. **25**, 253 (1977).
[8] C. W. Jurgensen and E. S. G. Shaqfeh, J. Vac. Sci. Technol. B **7**, 1488 (1989).
[9] M. Dalvie, R. T. Farouki, and S. Hamaguchi, IEEE Trans. Electron. Devices **39**, 1090 (1992).
[10] The tangential velocity component c_t determines the motion of the end points of the curve if the curve has end points.
[11] P. D. Lax, *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves*, SIAM Regional Conference Series in Applied Mathematics (SIAM, Philadelphia, 1973).
[12] R. E. Jewett, P. I. Hagouel, A. R. Neureuther, and T. van Duzer, Polym. Eng. Sci. **17**, 381 (1977).
[13] W. Fichtner, in *VSLI Technology*, edited by S. M. Sze (McGraw-Hill, New York, 1988).
[14] F. John, *Partial Differential Equations* (Springer, New York, 1982).
[15] J. Smoller, *Shock Waves and Reaction-Diffusion Equations* (Springer, New York, 1983).
[16] O. A. Oleinik, Ups. Mat. Nauk. **12**, 3 (1957); English transl. in Am. Math. Soc. Transl. Ser. 2 **26**, 95 (1963).
[17] M. G. Crandall and P. L. Lions, Trans. Am. Math. Soc. **277**, 1 (1983).
[18] M. G. Crandall, L. C. Evans, and P. L. Lions, Trans. Am. Math. Soc. **282**, 487 (1984).
[19] M. G. Crandall and P. L. Lions, Math. Comput. **43**, 1 (1984).